

HYPER TEXT ANALYZER, METHOD THEREFOR AND STORAGE MEDIUM RECORDED WITH HYPERTEXT ANALYTIC PROGRAM THEREIN

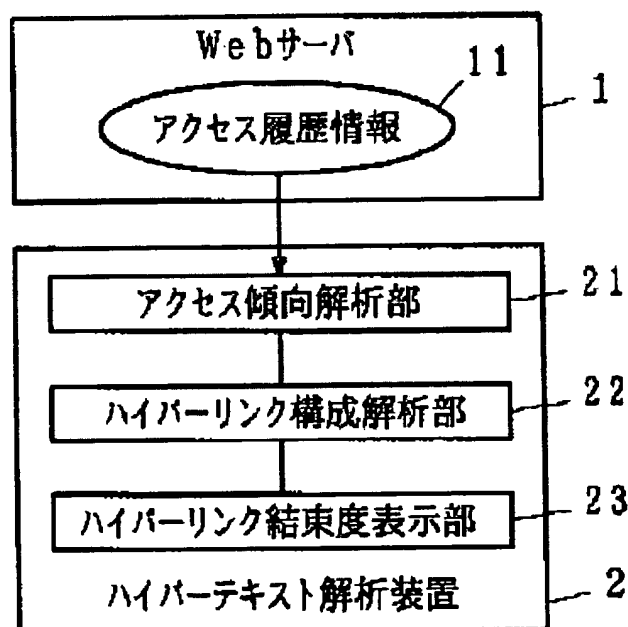
Patent number: JP2000172665
Publication date: 2000-06-23
Inventor: NAKAYAMA TAKEHIRO; KATO HIROKI
Applicant: FUJI XEROX CO LTD
Classification:
- international: G06F15/18; G06F12/00; G06F13/00; G06F15/00;
G06F17/30; G06N3/00; G06F15/18; G06F12/00;
G06F13/00; G06F15/00; G06F17/30; G06N3/00; (IPC1-
7): G06F15/18; G06F12/00; G06F13/00; G06F17/30
- european:
Application number: JP19980345759 19981204
Priority number(s): JP19980345759 19981204

Report a data error here

Abstract of JP2000172665

PROBLEM TO BE SOLVED: To provide a hypertext analyzer capable of recognizing the access tendency of a user from an access history and supporting the acquisition of knowledge for deciding superiority or inferiority in the configuration of a hypertext system.

SOLUTION: An access tendency analytic part 21 performs processing of clustering to a node group on a web server 1, while using respectively correspondent access history information 11. With respect to respective clusters generated by the access tendency analyzing part 21, a hyperlink configuration analyzing part 22 calculates the degree of hyper link coupling between nodes composing of the cluster. Since the result clustered based on the access history is used, this hyperlink coupling degree is reflected with the access tendency of the user. The value of the hyperlink coupling degree calculated by this hyperlink configuration analyzing part 22 is displayed on a hyperlink coupling degree display part 23 in a mode which is easy to utilize.



Data supplied from the esp@cenet database - Worldwide

THIS PAGE BLANK (USPTO)

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2000-172665

(P2000-172665A)

(43)公開日 平成12年6月23日(2000.6.23)

(51)Int.Cl. ⁷	識別記号	F I	テマコード(参考)
G 0 6 F 15/18	5 6 0	G 0 6 F 15/18	5 6 0 A 5 B 0 7 5
12/00	5 4 7	12/00	5 4 7 H 5 B 0 8 2
13/00	3 5 4	13/00	3 5 4 D 5 B 0 8 9
17/30		15/40	3 1 0 F
		15/419	3 2 0
審査請求 未請求 請求項の数6 O L (全 8 頁)			

(21)出願番号 特願平10-345759

(22)出願日 平成10年12月4日(1998.12.4)

(71)出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂二丁目17番22号

(72)発明者 中山 雄大

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(72)発明者 加藤 裕樹

神奈川県足柄上郡中井町境430 グリーン

テクなかい 富士ゼロックス株式会社内

(74)代理人 100101948

弁理士 柳澤 正夫

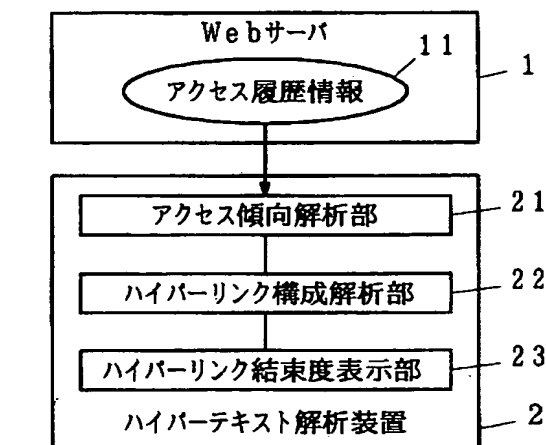
最終頁に続く

(54)【発明の名称】 ハイパーテキスト解析装置及び方法、ハイパーテキスト解析プログラムを記録した記憶媒体

(57)【要約】

【課題】 アクセス履歴からユーザのアクセス傾向を認識できるとともに、ハイパーテキストシステムの構成の優劣を判断するような知識を得ることを支援するハイパーテキスト解析装置を提供する。

【解決手段】 アクセス傾向解析部21は、Webサーバ1上のノード群に対し、それぞれに対応するアクセス履歴情報11を用いてクラスタリングの処理を行う。ハイパーリンク構成解析部22は、アクセス傾向解析部21で生成された各クラスタについて、クラスタを構成するノード間のハイパーリンク結束度を計算する。このハイパーリンク結束度は、アクセス履歴に基づいてクラスタリングされた結果を用いているので、ユーザのアクセス傾向を反映している。このハイパーリンク構成解析部22で計算されたハイパーリンク結束度の値を、ハイパーリンク結束度表示部23において利用しやすい形態で表示する。



【特許請求の範囲】

【請求項1】 ハイパーテキストシステムへのアクセス履歴情報に基づいて該ハイパーテキストシステムを構成するノードに対してクラスタリングを行うアクセス傾向解析手段と、該アクセス傾向解析手段で得られた各クラスタについて該クラスタを構成するノード間のハイパーリンク結束度を計算するハイパーリンク構成解析手段と、該ハイパーリンク構成解析手段で得られたハイパーリンク結束度を表示するハイパーリンク結束度表示手段を有することを特徴とするハイパーテキスト解析装置。

【請求項2】 前記ハイパーテキストシステムを構成する各ノード毎に予め定めたある一定期間にアクセスしてきたコンピュータの識別子を重複を許してリストとして保持する識別子獲得手段を有し、前記アクセス傾向解析手段は、該識別子獲得手段が保持しているリストから識別子の重複度合いを求めてクラスタリングを行うことを特徴とする請求項1に記載のハイパーテキスト解析装置。

【請求項3】 さらに、前記ハイパーテキストシステムを網羅的にアクセスして自動的に情報を収集しているコンピュータの識別子を特定し該識別子を前記識別子獲得手段が保持する前記リストから削除する識別子整形手段を有することを特徴とする請求項2に記載のハイパーテキスト解析装置。

【請求項4】 前記ハイパーリンク結束度が予め定めたある条件を満たすか否かを判定するハイパーリンク結束度評価手段を有し、前記ハイパーリンク結束度表示手段は、該ハイパーリンク結束度評価手段における判定結果に従ってハイパーリンク結束度を表示することを特徴とする請求項1ないし請求項3のいずれか1項に記載のハイパーテキスト解析装置。

【請求項5】 ハイパーテキストシステムへのアクセス履歴情報に基づいて該ハイパーテキストシステムを構成するノードに対してクラスタリングを行い、得られた各クラスタについて該クラスタを構成するノード間のハイパーリンク結束度を計算し、計算したハイパーリンク結束度を表示することを特徴とするハイパーテキスト解析方法。

【請求項6】 ハイパーテキストシステムへのアクセス履歴情報に基づいて該ハイパーテキストシステムを構成するノードに対してクラスタリングを行うアクセス傾向解析処理と、該アクセス傾向解析処理で得られた各クラスタについて該クラスタを構成するノード間のハイパーリンク結束度を計算するハイパーリンク構成解析処理と、該ハイパーリンク構成解析処理で得られたハイパーリンク結束度を表示してユーザに該ハイパーテキストシステムの構成の優劣を示すハイパーリンク結束度表示処理をコンピュータに実行させるためのハイパーテキスト解析プログラムを記録した記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワーク上に構成されるハイパーテキストシステムにおいて、その構成の優劣を判断するための知識を発見するためにハイパーリンク構造を解析するハイパーテキスト解析装置及びハイパーテキスト解析方法と、そのハイパーテキスト解析装置または方法をコンピュータで実現するためのハイパーテキスト解析プログラムを記録した記録媒体に関するものである。

【0002】

【従来の技術】ネットワーク上に構成されるハイパーテキストシステム（例えば、World Wide Web：以降Webと略す）では、ハイパーテキストを格納しているサーバにおいてユーザ（訪問者）のアクセス履歴を記録することができる。このアクセス履歴には、一般に、アクセスしてきたユーザが使用しているコンピュータの識別子（インターネットを利用しているのであればIPアドレス）、アクセスしてきた時刻、アクセスしたノードのサーバ上での識別子（WebではURL）が含まれる。

【0003】アクセス履歴を解析して、個々のユーザが遷移したハイパーリンクの経路を特定するには、基本的には、各コンピュータ毎にアクセスしたノードを時刻順に並べればよい。しかし、コンピュータのキャッシュ機能等により完全な経路を特定することは困難であった。

【0004】これに対し、例えば、C. Shahab i, A. M. Zarkesh, J. Adibi, and V. Shah, "Knowledge Discovery from Users Web-Page Navigation", in Proc. of IEEE RIDE, 1997. では、リモートエージェントを使うことによりキャッシュへのアクセスを認識している。これによって、より正確な経路を得ることができるようになった。ただし、ユーザのコンピュータ側でリモートエージェントプログラムをロードするためのコスト（時間とスペース）が犠牲となるという不具合がある。

【0005】アクセス履歴を利用して重要経路を発見する技術としては、例えば、J. Borges and M. Levene, "Mining Association Rules in Hypertext Databases", in Proc. of KDD, 1998. に記載されている技術がある。この技術は、まずアクセス履歴を収集して、ハイパーリンクのトラフィック量を重みとする有向グラフでハイパー構造を表現する。この有向グラフにおいて、ノードAからノードBへの遷移をA→Bと記し、これを結合規則と呼ぶ。結合規則は、コンフィデンス値（＝Aを起点とする遷移の総数に対するA→Bの遷移の総数）とサポート値（＝有向グラフ中のすべてのアークの遷移数の平均値に対するA→Bの遷移の総数）によって評価される。さらに、結合

10

20

30

40

50

規則を合成した合成結合規則 $(A \rightarrow B) \& (B \rightarrow C) \& (C \rightarrow D) \& \dots$ を定義して3つ以上のノードの遷移を評価している。

【0006】この手法では、コンピュータ識別子（IPアドレス）の情報は利用していないので、合成結合規則においては、単にトラフィック量が多いハイパーリンクの組み合わせ経路を発見しているに過ぎない。例えば、 $(A \rightarrow B) \& (B \rightarrow C)$ という合成結合規則が、高いコンフィデンス値とサポート値を持っていることがわかったとしても、実際に、 $A \rightarrow B \rightarrow C$ という経路を辿ったユーザが多かったとは限らない。

【0007】このように従来技術によって、ユーザのアクセス経路を特定したり、重要経路を発見することは可能である。しかし、ハイパーテキストシステム（例えば、Webサイト）の構成の優劣を判断するような知識を得ることはできなかった。

【0008】

【発明が解決しようとする課題】本発明は、上述した事情に鑑みてなされたもので、アクセス履歴からユーザのアクセス傾向を認識できるとともに、ハイパーテキストシステムの構成の優劣を判断するような知識を得ることを支援するハイパーテキスト解析装置及びハイパーテキスト解析方法を提供することを目的とするものである。また、そのハイパーテキスト解析装置または方法をコンピュータで実現するためのハイパーテキスト解析プログラムを記録した記録媒体を提供することを目的とするものである。

【0009】

【課題を解決するための手段】本発明は、ハイパーテキストシステムへのアクセス履歴情報に基づいて該ハイパーテキストシステムを構成するノードに対してクラスタリングを行い、得られた各クラスタについて該クラスタを構成するノード間のハイパーリンク結束度を計算し、計算したハイパーリンク結束度を表示することを特徴とするものである。表示されるハイパーリンク結束度は、ユーザがアクセスした履歴に基づいた値であるから、ユーザのアクセス傾向を示している。そのため、ハイパーリンク結束度を得ることによって、例えばハイパーテキストシステム（例えばWebサイト）のハイパーリンク構成などととともに、ハイパーテキストシステムの構成の優劣を判断することが可能となる。

【0010】

【発明の実施の形態】図1は、本発明の第1の実施の形態を示す構成図、図2は、Webの一例の説明図である。図中、1はWebサーバ、2はハイパーテキスト解析装置、11はアクセス履歴情報、21はアクセス傾向解析部、22はハイパーリンク構成解析部、23はハイパーリンク結束度表示部である。ネットワーク上に構成されるハイパーテキストシステムのなかで代表的なものはWebである。以下、Webを例として説明する。

【0011】Webサーバ1は、ネットワーク上で情報を発信する手段である。Webサーバ1には、図2に示すように、ユーザに提供したい情報がノード（矩形で示す）とハイパーリンク（矢線で示す）によるハイパー構造で貯えられている。ユーザはWebサーバ1にアクセスすることでコンテンツを入手できる。このとき、Webサーバ1では、一般に、ユーザのアクセスがある毎に、ユーザのコンピュータを識別するためのコンピュータ識別子（IPアドレス）とアクセス時刻とユーザのアクセスしたノードのあるアドレス（URL）をアクセス履歴情報11として記録している。

【0012】ハイパーテキスト解析装置2は、アクセス傾向解析部21、ハイパーリンク構成解析部22、ハイパーリンク結束度表示部23などを有している。アクセス傾向解析部21は、Webサーバ1上のノード群に対し、それぞれに対応するアクセス履歴情報11を用いてクラスタリングの処理を施す。このクラスタリングの処理には既存の技術を使用することができる。例えば、Agglomerative Hierarchical Clusteringによるクラスタリングの方法では、以下の1、2、3のステップを行う。

1. Webサーバ1上の各ノードをそれぞれ一つのクラスタとする。
2. 各クラスタ間の類似度を計算し、最大類似度を持つクラスタ同士を一つのクラスタにマージする。
3. クラスタが一つになるまで2の処理を繰り返す。

この過程の途中で順次生成されるそれぞれのクラスタをクラスタリングの処理結果として得る。例えば1.の処理において生成される各ノードのみのクラスタ、そのクラスタをマージした各クラスタ、最後に生成された1つのクラスタなどがクラスタリング結果となる。なお、上述のクラスタリングの方法は、例えば、E. M. Voorhees, "Implementing Agglomerative Hierarchical Clustering Algorithms for Use in Document Retrieval", Information Processing & Management, Vol. 22, No. 6, 1986等に記載されている。

【0013】アクセス傾向解析部21では、上述のクラスタリング処理の2.の処理における類似度計算において、アクセス履歴情報11の重複度合いを利用する。具体的には、例えば、アクセス履歴情報11の中のIPアドレスに注目して、クラスタ間で共有するIPアドレスの数を類似度と定義する。または、クラスタ間で共有するIPアドレスの数を当該二つのクラスタが持つ全てのIPアドレスの数で割った値を類似度と定義する。ここで、クラスタが持つIPアドレスとは、クラスタを構成する各ノードへのアクセス履歴を全てマージして、その中に出現するIPアドレスのことである。IPアドレス

の数は、このような IP アドレスから重複を排除した異なる IP アドレスを計数した値とすることができる。なお、IP アドレスを用いる代わりに、例えばアクセス履歴情報 11 中のアクセス時刻を用いてもよい。この場合、秒や分単位ではなく日付や月の単位で取り扱うとよい。

【0014】あるいは、類似度計算として、各クラスタを該クラスタに対応するアクセス履歴情報 11 を項とし、該アクセス履歴情報の出現頻度を該項の値とするようなベクトルを生成して、ベクトル間の内積値の大小を類似度として用いてもよい。

【0015】ハイパーリンク構成解析部 22 は、アクセス傾向解析部 21 で生成された各クラスタについて、クラスタを構成するノード間のハイパーリンク結束度（以下、単に結束度と呼ぶ）を計算する。結束度は、例えば、ノード間に一つ以上のハイパーリンクが存在すれば、そのノード間には結合があると定義したときに、クラスタを構成するノード間結合の総数を、クラスタを構成する全てのノードから二つを選ぶ組み合わせ数で割った値とすることができる。すなわち、ノード間結合の総数を L、ノード数を N としたとき、

$$\text{結束度} = L / N \cdot C_2$$

で計算することができる。ハイパーリンク構成解析部 22 で計算した、各クラスタの結束度の値は、ハイパーリンク結束度表示部 23 に渡される。

【0016】ハイパーリンク結束度表示部 23 は、ハイパーリンク構成解析部 22 で計算された結束度の値を、利用しやすい形態で表示する。図 3 は、ハイパーリンク結束度表示部による表示の一例の説明図である。例えば、結束度の値をいくつかのセグメントに分け、各セグメントに入るクラスタの数を棒グラフで表示することができる。図 3 では、クラスタ数の分布を縦軸とし、結束度を横軸として、二つの Web サイト（Web サイト A と Web サイト B）について並べて表示したものである。前述のように、各クラスタはアクセス履歴情報を基に構成されているので、各クラスタを構成するノード群は、例えば、同一のユーザから前後してアクセスされる傾向が強いものである。ノード群の結束度が高いと、ユーザにとってはノード間遷移のための経路が多数提供されることになるので効率よくブラウジングできることになる。一方、該ノード群の結束度が低いとブラウジング効率は悪くなる。図 3 では、Web サイト A の方は結束度が低いクラスタが多数あり、Web サイト B の方は結束度が高いクラスタが多数あるということが一目でわかる。この表示によって、Web サイト B の方が Web サイト A よりも優れた構成でハイパーテキストシステムが構築されていると容易に判断できる。

【0017】ハイパーリンク結束度表示部 23 は、図 3 に示す表示形態のほか、各種の表示形態により結束度を表示することが可能である。図 4 は、ハイパーリンク結

束度表示部による表示の別の例の説明図である。図 4 に示した例では、クラスタのサイズと結束度の値の関係を表示している。図中の点は、それぞれがクラスタを示している。クラスタのサイズとしては、例えば、クラスタを構成するノード数や、クラスタを構成する各ノードが持つ単語の総数や、クラスタを構成する各ノードのファイルサイズの総計などを用いることができる。図 4 に示した例では、クラスタのサイズとしてクラスタを構成するノード数を用いて表示した例を示している。このような表示を行った場合、同じクラスタサイズであれば、結束度が高いほど優れた構成であると判断することができる。また、このような表示によって、クラスタサイズに注目しながら、各クラスタの構成の優劣を俯瞰することができる。

【0018】図 5 は、ハイパーリンク結束度表示部による表示の別の例においてクラスタを選択した場合の表示例の説明図である。図 4 に示したような結束度の表示が行われているとき、クラスタを表す点をマウス等のポインティングデバイスで選択すると、図 5 に示すように、選択されたクラスタを構成するノードの URL（識別子）を表示できるように構成することができる。あるいは、URL ではなく、ノードのタイトルを表示してもよい。さらに、図 5 に示すように表示されたノード（URL で表示されている）をマウス等のポインティングデバイスで選択すると、ネットワークを通じて選択したノードにアクセスして、そのノードのコンテンツを獲得し、そのコンテンツを表示するようにしてもよい。このような表示によって、Web サイトの管理者は、Web サイト内の問題箇所へアクセスしてコンテンツを参照することができ、さらに編集することができるので、Web サイト内の構成を容易に改善して行くことができる。

【0019】図 6 は、ハイパーリンク結束度表示部による表示の別の例においてクラスタを選択した場合の別の表示例の説明図である。図 5 に示した例と同様に、図 4 に示したような結束度の表示が行われているとき、クラスタを表す点をマウス等のポインティングデバイスで選択することにより、図 6 に示すように選択したクラスタを構成するノードとそのノード間のハイパーリンクを表示することもできる。ここで、ノードのラベル（図 6 中では、N1、N2、…、N10）は、URL でもタイトルでもよい。このような表示を行うことによって、Web サイトの管理者は、Web サイト内の問題箇所を容易に発見することができる。例えば図 6 に示したクラスタ内のハイパーリンクの表示例では、N1～N7 と N8～N10 の間にハイパーリンクがないのでユーザは両グループ間を容易に行き来できないことがわかる。

【0020】図 7 は、ハイパーリンク結束度表示部による表示のさらに別の例の説明図である。この例では、クラスタ内のノード間の類似度と結束度の値との関係を表示した例を示している。クラスタ内のノード間の類似度

10

20

30

40

50

としては、例えばアクセス傾向解析部 21 においてクラスタ生成時に用いた類似度の値を用いることができる。このような表示において、同じクラスタ内類似度であれば、結束度が高いほど優れた構成であると判断することができる。このような表示によって、クラスタ内のノード間の類似度に注目しながら、各クラスタの構成の優劣を俯瞰することができる。

【0021】なお、ハイパーリンク結束度表示部 23 では、上述の各例に示した表示形態によらず、任意の形態で結束度を表示させることができる。

【0022】図 8 は、本発明の第 2 の実施の形態を示す構成図である。図中、図 1 と同様の部分には同じ符号を付して説明を省略する。24 は識別子獲得部である。識別子獲得部 24 は、アクセス履歴情報 11 の中から Web サーバ 1 上の各ノード毎に予め定められた一定期間にアクセスしてきたコンピュータの識別子（例えば IP アドレス）を獲得する。

【0023】アクセス傾向解析部 21 は、第 1 の実施の形態と同様に、Web サーバ 1 上のノード群に対し、それぞれに対応するアクセス履歴情報 11 を用いてクラスタリングの処理を施す。このクラスタリングの処理において類似度を計算する際に、識別子獲得部 24 において獲得したコンピュータの識別子の重複度合いを利用することができる。

【0024】図 9 は、識別子獲得部における処理の一例を示すフローチャートである。まず、S31 において、アクセス履歴情報 11 の中から、予め定められた期間内のアクセスに関するものだけを抽出し、残りを破棄する。S32 において、S31 で抽出されたアクセス履歴情報 11 に存在する全てのコンピュータ識別子（IP アドレス）のうちの異なる識別子の数（異なり数）N を求める。次に S33 において、アクセス履歴情報 11 に存在する全てのノードについて、各ノード毎に次元数 N のベクトルを生成する。ここで、ベクトルの各項目は、アルファベットの昇順（あるいは降順）に並べた互いに異なるコンピュータ識別子（IP アドレス）に対応する。各項目の初期値は 0 としておく。次に S34 において、アクセス履歴情報 11 の中から各ノードにアクセスしてきたコンピュータのコンピュータ識別子（IP アドレス）を取り出し、対応するベクトルの項に 1 を加える。この処理を、アクセス履歴情報 11 すべてについて、順に行う。

【0025】このようにして得られた N 次元のベクトルを用いて、アクセス傾向解析部 21 では、ベクトル間の内積値の大小を類似度としてクラスタリングをすることができる。ここでクラスタとクラスタをマージする際は、それぞれのベクトルの和を取れば、これがマージされたクラスタのベクトルとなる。

【0026】以降の処理は上述の第 1 の実施の形態と同様である。ハイパーリンク構成解析部 22 において結束

度を計算し、ハイパーリンク結束度表示部 23 において結束度を表示する。このとき、上述のような各種の表示形態あるいはそれ以外の各種の表示形態で結束度を表示することができる。

【0027】図 10 は、本発明の第 3 の実施の形態を示す構成図である。図中、図 1 と同様の部分には同じ符号を付して説明を省略する。25 はコンピュータ識別子整形部である。コンピュータ識別子整形部 25 は、Web サーバ 1 に対するアクセスのうち、ユーザ（＝人間）によって操作されるコンピュータからのものではなく、Web を網羅的にアクセスして自動的に情報を収集している情報収集ロボットのようなコンピュータによるアクセスを排除する。例えば、ある慣習に従って情報収集ロボットがアクセスする特殊なノード（Web では、例えばルート直下に置かれる robots.txt という名のファイル）へのアクセスの有無によって、情報収集ロボットからのアクセスであるか否かを判断することができる。あるいは、短期間に多数のノードを網羅的にアクセスするという情報収集ロボットに特徴的な振る舞いの有無や、既知の情報収集ロボットのコンピュータ識別子であるか否かによっても、情報収集ロボットを識別することができる。情報収集ロボットからのアクセスであると判断されたコンピュータについては、そのコンピュータ識別子に関わるアクセス履歴情報 11 を、識別子獲得部 24 において獲得しないようにすることができる。

【0028】これによって、情報収集ロボットのような Web を網羅的にアクセスして自動的に情報を収集しているコンピュータによるアクセスを排除し、解析結果に対するこれらの影響を除去することができ、ユーザのアクセス動向を正しく反映した解析結果を得ることができる。

【0029】なお、この第 3 の実施の形態におけるその他の構成および動作は、上述の第 2 の実施の形態と同様である。

【0030】図 11 は、本発明の第 4 の実施の形態を示す構成図である。図中、図 1 と同様の部分には同じ符号を付して説明を省略する。26 はハイパーリンク結束度評価部である。ハイパーリンク結束度評価部 26 は、ハイパーリンク構成解析部 22 で得られた各クラスタの結束度の値を予め定められた閾値と比較し、結束度が閾値よりも小さなクラスタをハイパーリンク結束度表示部 23 に渡す。これによって、ハイパーリンク結束度表示部 23 では結束度が小さい、すなわち構成が劣るクラスタとそのクラスタを構成するノードを容易に得ることができる。

【0031】図 12 は、ハイパーリンク結束度表示部による表示のさらに別の例の説明図である。図 12 に示した表示例では、結束度の値が小さい順にクラスタおよびそのクラスタを構成するノードの識別子（URL）を表示している。クラスタは「ID」の欄に示している。こ

こで「ID」は、クラスタを参照するためにユニークにつけられた番号である。例えば、アクセス傾向解析部21におけるクラスタ生成時に、生成された順に番号を付与すればよい。このような表示によって、構成上劣っている部分から表示されるので、ユーザが利用する上でネックとなっている部分を容易に知ることができる。もちろん、この第4の実施の形態においても、第1の実施の形態で示したような各種の表示形態あるいはその他の表示形態で結束度を表示することが可能である。また、第2、第3の実施の形態で説明した識別子獲得部24、コンピュータ識別子政経部25などを設けてもよい。

【0032】上述の実施の形態は、コンピュータプログラムによっても実現することが可能である。その場合、そのプログラムおよびそのプログラムが用いるデータなどは、コンピュータが読み取り可能な記憶媒体に記録しておくことも可能である。記憶媒体とは、コンピュータのハードウェア資源に備えられている読取装置に対して、プログラムの記述内容に応じて、磁気、光、電気等のエネルギーの変化状態を引き起こして、それに対応する信号の形式で、読取装置にプログラムの記述内容を伝達できるものである。例えば、磁気ディスク、光ディスク、CD-ROM、コンピュータに内蔵されるメモリ等である。

【0033】

【発明の効果】以上の説明から明らかなように、本発明によれば、ユーザのアクセス履歴情報に基づいて、ハイパーテキストシステムを構成するノードに対してクラスタリングを行い、得られた各クラスタを構成するノード間のハイパーリンク結束度を計算する。このハイパーリンク結束度によって、ハイパーテキストシステムの構成の優劣を容易に判断することが可能になる。例えばWebの管理者は、構成に問題がある部分に変更を加えて、より良い構成のハイパーテキストシステムを構築するこ*

*とができるという効果がある。

【図面の簡単な説明】

【図1】 本発明の第1の実施の形態を示す構成図である。

【図2】 Webの一例の説明図である。

【図3】 ハイパーリンク結束度表示部による表示の一例の説明図である。

【図4】 ハイパーリンク結束度表示部による表示の別の例の説明図である。

10 【図5】 ハイパーリンク結束度表示部による表示の別の例においてクラスタを選択した場合の表示例の説明図である。

【図6】 ハイパーリンク結束度表示部による表示の別の例においてクラスタを選択した場合の別の表示例の説明図である。

【図7】 ハイパーリンク結束度表示部による表示のさらに別の例の説明図である。

【図8】 本発明の第2の実施の形態を示す構成図である。

20 【図9】 識別子獲得部における処理の一例を示すフローチャートである。

【図10】 本発明の第3の実施の形態を示す構成図である。

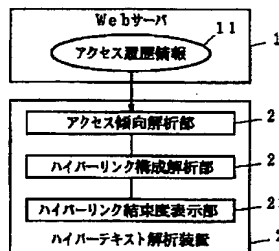
【図11】 本発明の第4の実施の形態を示す構成図である。

【図12】 ハイパーリンク結束度表示部による表示のさらに別の例の説明図である。

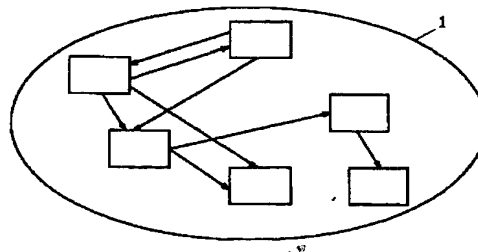
【符号の説明】

1…Webサーバ、2…ハイパーテキスト解析装置、11…アクセス履歴情報、21…アクセス傾向解析部、22…ハイパーリンク構成解析部、23…ハイパーリンク結束度表示部、24…識別子獲得部、25…コンピュータ識別子整形部、26…ハイパーリンク結束度評価部。

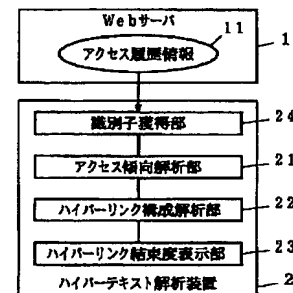
【図1】



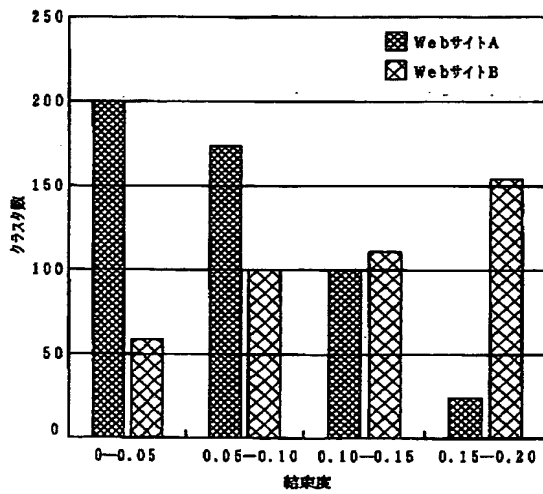
【図2】



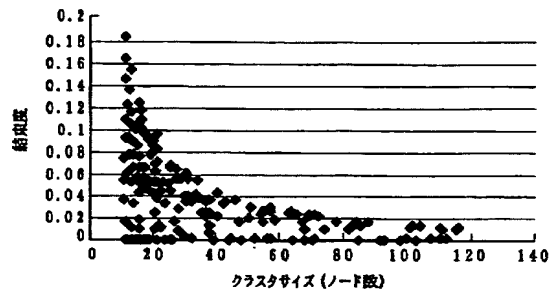
【図8】



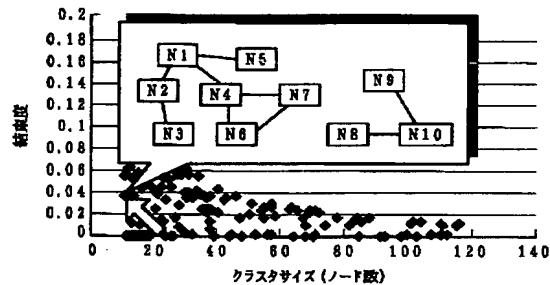
【図3】



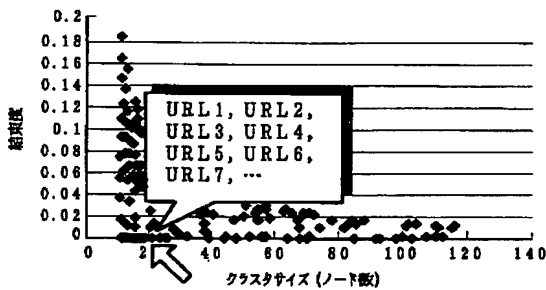
【図4】



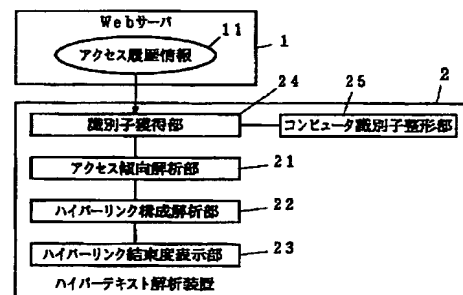
【図6】



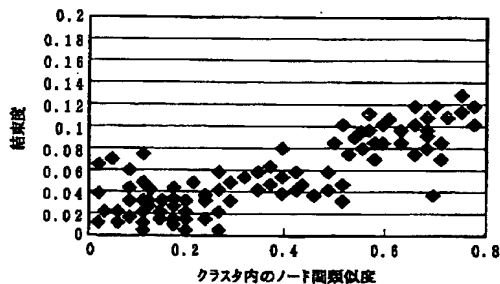
【図5】



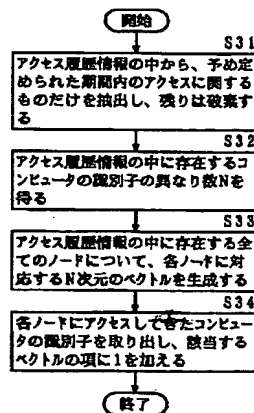
【図10】



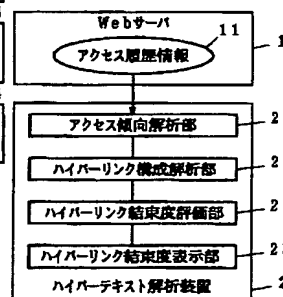
【図7】



【図9】



【図11】



【図12】

結束度	ID	構成ノード
0.01	#101	URL1, URL2, URL3, URL4, URL5, URL6
0.03	#211	URL11, URL12, URL13, URL14, URL15
0.04	#315	URL21, URL22, URL23, URL24
0.06	#118	URL31, URL32, URL33, URL34, URL35
0.08	#56	URL41, URL42, URL43

フロントページの続き

Fターム(参考) 5B075 ND20 NR12 PQ02 PQ14 PQ36
PR03
5B082 AA00 EA07 FA11 GC03 HA08
5B089 GA11 GB02 HA01 JA21 JA36
JB02 JB11 KA13 KA14 KB06
KC36 LB14